16. Pollitz, F. F. Transient rheology of the uppermost mantle beneath the Mojave Desert, California. *Earth Planet. Sci. Lett.* **215,** 89–104 (2003).
17. Wald, D. J. & Heaton, T. H. Spatial and temporal distribution of slip for the 1992 Landers, California, earthquake. *Bull. Seismol. Soc. Am.* **84,** 668–691 (1994).
18. Kaverina, A., Dreger, D. & Price, E. The combined inversion of seismic and geodetic data for the source process of the 16 October 1999 Mw 7.1 Hector Mine, California, earthquake. *Bull. Seismol. Soc. Am.* **92,** 1266–1280 (2002).
19. Masterlark, T. & Wang, H. F. Transient stress-coupling between the 1992 Landers and 1999 Hector Mine, California, earthquakes. *Bull. Seismol. Am.* **92,** 1470–1486 (2002).
20. Peltzer, G., Rosen, P. & Rogez, F. Poroelastic rebound along the Landers 1992 earthquake surface rupture. *J. Geophys. Res.* **103,** 30131–30145 (1998).
21. Williams, C. F. Temperature and the seismic/aseismic transition; observations from the 1992 Landers earthquake. *Geophys. Res. Lett.* **23,** 2029–2032 (1996).
22. Melbourne, T. & Helmberger, D. Mantle control of plate boundary deformation. *Geophys. Res. Lett.* **28,** 4003–4006 (2001).
23. Goes, S. & van der Lee, S. Thermal structure of the North American uppermost mantle inferred from seismic tomography. *J. Geophys. Res.* **107,** doi:2000JB000049 (2002).
24. Farmer, G. L. *et al.* Origin of late Cenozoic basalts at the Cima volcanic field, Mojave Desert, California. *J. Geophys. Res.* **100,** 8399–8415 (1995).
25. Kronenberg, A. K. & Tullis, J. A. Flow strengths of quartz aggregates; grain size and pressure effects due to hydrolytic weakening. *J. Geophys. Res.* **89,** 4281–4297 (1984).
26. Shelton, G. & Tullis, J. A. Experimental flow laws for crustal rocks. *Trans. Am. Geophys. Union* **62,** 396 (1981).
27. Jaoul, O., Tullis, J. A. & Kronenberg, A. K. The effect of varying water contents on the creep behaviour of Heavitree Quartzite. *J. Geophys. Res.* **89,** 4298–4312 (1984).
28. Hansen, F. D. & Carter, N. L. Creep of selected crustal rocks at 1000 MPa. *Trans. Am. Geophys. Union* **63,** 437 (1982).
29. Hirth, G. & Kohlstedt, D. Rheology of the upper mantle and the mantle wedge: A view from the experimentalists. In *The Subduction Factory* (ed. Eiler, J.) (American Geophysical Union 2004).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements Funding was provided by the National Science Foundation and the Southern California Earthquake Center. GPS data were provided by the US Geological Survey and the Southern California Integrated GPS Network..

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.M.F. (freed@purdue.edu).

Fine-scale phylogenetic architecture of a complex bacterial community

Silvia G. Acinas[1]*, Vanja Klepac-Ceraj[1]*, Dana E. Hunt[1], Chanathip Pharino[1], Ivica Ceraj[2], Daniel L. Distel[3] & Martin F. Polz[1]

[1]*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[2]*Bugaco, Somerville, Massachusetts 02144, USA*
[3]*Department of Biochemistry, Microbiology and Molecular Biology, University of Maine, Orono, Maine 04469, USA*

* These authors contributed equally to this work

Although molecular data have revealed the vast scope of microbial diversity[1], two fundamental questions remain unanswered even for well-defined natural microbial communities: how many bacterial types co-exist, and are such types naturally organized into phylogenetically discrete units of potential ecological significance? It has been argued that without such information, the environmental function, population biology and biogeography of microorganisms cannot be rigorously explored[2]. Here we address these questions by comprehensive sampling of two large 16S ribosomal RNA clone libraries from a coastal bacterioplankton community. We show that compensation for artefacts generated by common library construction techniques reveals fine-scale patterns of community composition. At least 516 ribotypes (unique rRNA sequences) were detected in the sample and, by statistical extrapolation, at least 1,633 co-existing ribotypes in the sampled population. More than 50% of the ribotypes fall into discrete clusters containing less than 1% sequence divergence. This pattern cannot be accounted for by interoperon variation, indicating a large predominance of closely related taxa in this community. We propose that such microdiverse clusters arise by selective sweeps and persist because competitive mechanisms are too weak to purge diversity from within them.

Traditional species concepts have largely been concessions to the need to identify bacteria reproducibly, but none adequately describe natural units of microbial diversity[3]. It has recently been proposed that natural taxa are distinct groups of strains that arise by periodic selection—a process of continuing, selectionally neutral, diversification punctuated by adaptive mutations leading to selective sweeps[4]. The latter events purge all sequence variants except those associated with the genome carrying the adaptive mutation[4]. One of the attractive features of this concept is that it should be applicable to molecular surveys of microbial diversity because taxa would be identifiable in phylogenetic trees as distinct clusters of closely related sequences[1]. Moreover, such clusters should be detectable independently of the gene used to construct these trees as long as the accumulation of variation is commensurate with the occurrence of sweeps[5]. However, this theory has not been applied to broad-scale studies of bacterial diversity in the environment. Over the past 20 years, diversity studies have primarily been based on analyses of 16S rRNA clone libraries but it has remained uncertain to what extent fine-scale patterns of variation are due to sequence artefacts, to heterogeneity among paralogous operons or to the co-existence of similar but differentiated taxa[1]. Furthermore, it has not been explored whether naturally defined units of differentiation emerge from recently released shotgun sequence data from the Sargasso Sea[6].

We deduced that the discovery of ecologically significant patterns of relationships between co-existing ribotypes requires, first, an examination of clone libraries large enough to elucidate relationships at all levels of differentiation, and second, methods that minimize and account for the contribution of sequence artefacts and paralogous variation to diversity estimates. We sequenced about 1,000 clones from each of two polymerase chain reaction (PCR)-derived 16S rRNA libraries constructed from the same coastal bacterioplankton sample. The first library employed common (standard) amplification protocols. For the second, a modified protocol was designed to minimize artefacts and to identify Taq errors and chimaeric molecules through extensive sequence analyses (see Methods). This approach allowed the most comprehensive analysis of any single gene from co-occurring populations so far, even in view of the recently released Sargasso Sea study, which in aggregate sampled a similar number of rRNA genes but from several locations, dates and diverse biogeochemical conditions[6]. Our overall rationale was to achieve high coverage of rRNA genes from a single community while estimating and compensating for the influence of artefacts on ribotype diversity, potentially revealing emergent patterns.

Comparison of the two libraries showed that changes to the amplification protocol alone decreased the incidence of unique sequences from 76% (692 of 909) in the standard to 61% (686 of 1,131) in the modified library. Correction for chimaeras and Taq error lowered the percentage to 48% (516 of 1,067) unique sequences (Fig. 1a), demonstrating a potentially significant contribution of PCR-induced artefacts to (micro)diversity estimates. Consequently, these corrections yield a significantly lower estimate of total ribotype diversity for the sampled community when compared with the unmodified standard library (1,633 versus 3,881) with the use of the Chao-1 estimator[7]. A novel estimator (N_T/N_{max}) (ref. 2) yielded a similar value of 2,236 sequences for the corrected data set. This good agreement, combined with the low incidence of chimaeras and the observation that corrections account

for most expected Taq errors (see Methods), provides confidence in the corrected estimate of ribotype diversity.

A vast and previously unrecognized predominance of microdiverse ribotypes was revealed by further analysis of relationships. More than half of the observed sequences in the modified library fell into clusters sharing at least 99% sequence consensus (Fig. 1a). This result is still more marked when the Chao-1 diversity estimator is applied to the data, indicating that more than two-thirds of ribotypes might be members of 99% sequence clusters in the sampled bacterioplankton. Defining such 99% clusters, rather than unique ribotypes, as operational taxonomic units (OTUs) decreases diversity estimates from 1,633 to 520 OTUs, a decline of about 70%. However, further clustering into 98% and 97% consensus groups decreases the number of OTUs by only 3% (507 OTUs) and 11% (450 OTUs), respectively (Fig. 1a). In fact, a remarkably consistent exponential decline was observed in the number of OTUs as cluster cut-off values were decreased from 99% to 75% (Fig. 1b). In stark contrast, the number of OTUs greatly exceeds this exponential trend for values above 99% (Fig. 1b). An essentially identical relationship emerged from a phylogenetic (maximum-likelihood) analysis, in which the accumulation of lineages per arbitrary time unit was inferred under a molecular clock model[8] (data not shown). This exponential accumulation of clusters or lineages is expected if the creation and removal of taxa are on average constant over time[8]. The sharp discontinuity observed above 99% similarity therefore suggests increased diversification or decreased removal of diversity within microdiverse clusters.

The overall predominance of extremely closely related ribotypes also emerges from phylogenetic analyses as large clusters of closely related taxa (Fig. 2, and Supplementary Information). These are typically well separated from other clusters, as indicated by a comparison of average within-cluster and between-cluster sequence divergence (data not shown)[5]. The most sequence-rich microdiverse clusters are formed within the Pelagibacter (SAR11) group of the alpha-Proteobacteria (Fig. 2) but all highly represented lineages contain such clusters, including the gamma-Proteobacteria and the Bacteriodetes group (Supplementary Information). Nevertheless, microdiverse clusters are not uniformly distributed between lineages in the modified clone library. For example, the Cytophaga group contains more deeply divergent lineages and fewer microdiverse clusters than the Pelagibacter group (Supplementary Information). However, such differences might be due to incomplete sampling because rarefaction (Fig. 1a) suggests that deeper branching lineages in this library are well sampled and that additional sequencing should therefore primarily reveal microdiverse ribotypes.

To what extent can the observed ribotype microdiversity be explained by variation among paralogous operons within single genomes[1]? We have recently explored this question by an analysis of 97 available complete bacterial genomes[9]. These contain, because of multiple non-identical operons, a total of 242 different 16S rRNA sequences[9]; that is, the number of ribotypes exceeds the number of genomes about 2.5-fold. Remarkably, interoperon sequence difference remains within about 1% among these genomes. Only five genomes deviate from this rule, four of which were thermophilic bacteria all with a single operon with higher sequence divergence[9]. Therefore, if one accepts that the distribution of operons among free-living bacteria is similar to that of the 97 sequenced genomes, a conservative correction factor of about 2.5 (ref. 9) can be applied to

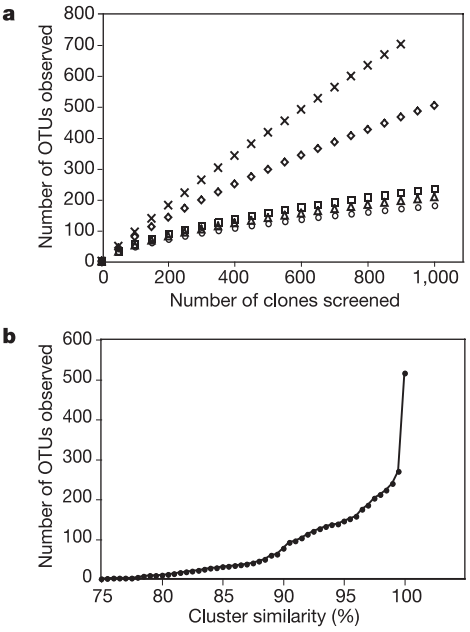

Figure 1 Compositional pattern of the coastal bacterioplankton sample. **a**, Rarefaction curves of the number of OTUs in a 16S rRNA library constructed with standard (crosses, 100% sequence similarity cluster) and modified (diamonds, 100% sequence similarity clusters; squares, 99%; triangles, 98%; circles, 97%) amplification protocols. Standard deviations fall within the symbols and are not shown. **b**, Number of OTUs plotted against changing degrees of cutoffs in 0.5% increments for grouping of sequences into similarity clusters.

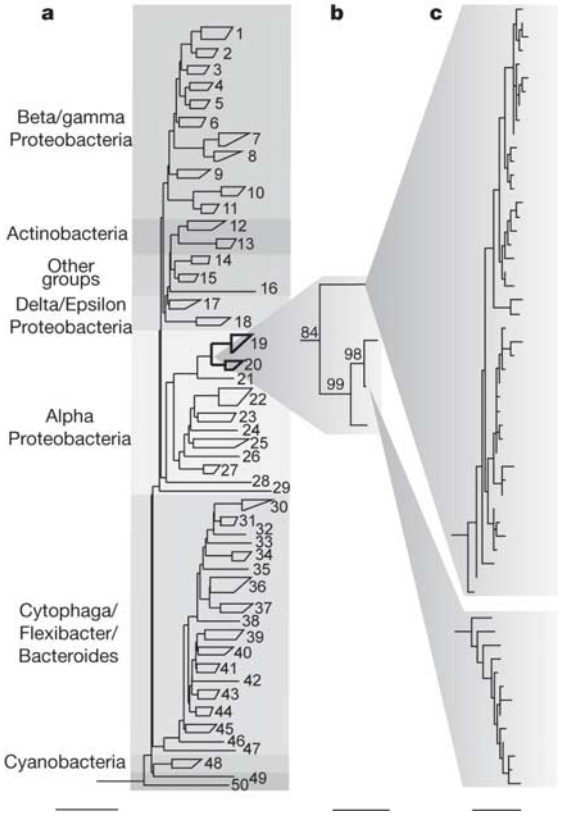

Figure 2 Phylogenetic distance relationships between the coastal bacterioplankton based on partial 16S rRNA sequencing. **a**, Summary of groups represented in the sample, in which each number denotes a phylogenetic cluster of sequences (for identification key see Supplementary Information). **b**, Relationships between *Pelagibacter* (SAR11) clusters represented by one sequence of each 99% similarity cluster. Numbers associated with nodes represent bootstrap values. **c**, Examples of microdiverse relationships between SAR11 ribotypes. Scale bars, 0.1 (**a**), 0.05 (**b**) and 0.01 (**c**) substitutions per site. Arrows connecting trees point to expanded nodes.

the estimated number of sequences (1,113) in the 99% similarity group to yield a revised estimate of at least 446 closely related genomes co-existing in the sample. However, this is probably an overcorrection because opportunistic bacteria with multiple operons are thought to predominate in culture collections and among sequenced genomes[9,10]. Moreover, *Pelagibacter ubique* HTCC1062, which is identical in sequence to clones within the largest SAR11 99% similarity clusters, seems to contain a single rRNA operon (S. J. Giovannoni, personal communication). Given that operon numbers vary little between closely related bacteria[9] it is unlikely that the observed SAR11 microdiversity can be explained by operon differences. Finally, shotgun sequencing of Sargasso Sea prokaryotes revealed a total of about 1,400 rRNA and about 600 RecA sequences[6]. The latter is a single-copy gene in all currently published genomes. Their frequency in the sample therefore provides an independent and almost identical estimate of 2.3 rRNA operons per genome. Thus, we conclude that, even after conservative correction, genomes denoted by microdiverse ribotypes represent by far the dominant fraction of bacterial diversity in this coastal bacterioplankton sample.

The observed pattern raises the question: what level of similarity should be expected between genomes carrying microvariant ribotypes? Comparative genomics has shown that genomes can be divided into stable and variable sets of genes, termed the core and flexible/auxiliary genome, respectively[11,12]. The latter arises primarily by means of phage and transposon-mediated lateral gene transfer and comprises between 1% and 18% of genes[11] but possibly as much as 60% (ref. 13). The core genome, in contrast, is a stable complement of genes that includes rRNA and housekeeping genes. This core reflects the overall evolutionary history of the lineage because little lateral gene transfer is detectable[9,12,14]. Microdiverse ribotype clusters should therefore also be apparent in comparisons of other housekeeping genes, possibly more so because of the higher substitution rates typical of protein coding genes[5].

Do microdiverse sequences denote co-existing, ecologically differentiated genomes? Among free-living bacteria of very similar ribotypes, correlation of genomic variation with ecological parameters has been demonstrated convincingly in only a single case involving two strains of *Prochlorococcus*[15], but these would not fall within a single microdiverse cluster as defined here. In contrast, no evidence of functional differentiation was detected in several environmental BAC clones with microdiverse 16S rRNA, despite considerable polymorphisms in protein-coding genes[16,17]. This is consistent with recently advanced theories for the interpretation of microdiverse sequence clusters. It has been shown[5] that clustering of housekeeping genes, resulting from periodic selection, predicts ecologically differentiated strains within cultivated bacterial taxa. If microdiverse ribotype clusters in the environment arise by the same mechanism[1] their very existence implies that intracluster competition is too weak to sweep members from within their ranks. However, this does not require that these genomes are functionally identical. Subdifferentiation within the flexible genome might provide increased fitness under episodic or spatially confined environmental conditions, but not sufficient growth advantage to sweep competing microdiverse genomes[12]. Furthermore, ecological factors might decrease effective competition. Particularly, predation has been suggested to promote the coexistence of diverse lineages by 'killing the winner' of competitive events[18]. Finally, recombination might be important in delineating and preserving genetic diversity among members of clusters by allowing sweeps of adaptive alleles without removing selectionally neutral variation[19].

The above considerations lead us to suggest that microdiverse ribotype clusters are important units of differentiation in natural bacterial communities. Indeed, such clusters might be widespread. We have recently detected numerous microdiverse ribotypes in salt-marsh sulphate-reducing bacteria[20], and ribotype clusters have previously been tentatively suggested for some open-ocean microbial groups[1]. To determine whether microdiverse ribotype clusters described here represent ecotypes—that is, ecologically cohesive populations—will require a detailed comparison of their encompassed genomic variation. Indeed, high-throughput sequencing[6] and cultivation[21] provide the means for rigorous testing of the hypothesized ecological importance of ribotype clusters. Most importantly, such inquiries challenge us to re-examine concepts of microbial diversity and invigorate the search for ecologically and evolutionarily defined species concepts. □

Methods

Study site and sampling

A 2.2-litre water sample was collected on 6 October 2001 from the marine end of the Plum Island Sound estuary (northeastern Massachusetts), and bacterioplankton was concentrated on a 0.22-μm filter (Supor; Gelman), which was stored at −80 °C until DNA extraction. Measured water parameters were as follows: temperature 16 °C, pH 8.0, prokaryotic cell numbers $0.99 \times 10^9 l^{-1}$, dissolved organic carbon 0.4 mg C l^{-1}, chlorophyll *a* 5.94 μg l^{-1}.

DNA extraction and clone library construction

Cells on filters were lysed and nucleic acids were extracted with a modified version of a bead-beating method[20] followed by treatment with RNase I and purification on Qiagen DNA purification spin columns. Two 16S rRNA clone libraries were constructed from the same sample to estimate the total coexisting sequence diversity and the effect of PCR-induced artefacts. For both, the bacteria-specific primers 27F and 1492R as modified in ref. 22 were used. Each PCR reaction contained genomic DNA equivalent to 4.9×10^6 cells. Ten replicate reactions were combined and gel-purified, and the same amount of amplicons were cloned with the PCR 2.1-TOPO kit (Invitrogen). The PCR amplification for the first (standard) library used 35 cycles mimicking commonly used protocols (typically between 30 and 40 cycles). The second (modified) library was constructed to minimize the accumulation of the three known PCR artefacts (Taq errors, chimaeras and heteroduplex molecules)[20]. In brief, the sample was amplified for 15 cycles followed by a 3-cycle 'reconditioning step', which eliminated heteroduplex molecules[23] and decreased the incidence of Taq errors and chimaeras. Purified plasmids served as templates for partial 16S rRNA sequence determination with the bacterial primer 27F (ref. 20).

Sequence analysis

Sequences (position 68 to 805, *E. coli* numbering) from the corrected library were further analysed for evidence of PCR artefacts. About 3% of the sequences were removed as putative chimaeras on the basis of identification by a combination of the three bioinformatics tools Chimera_Check[24], Bellerophon[25] and ChimeraBuster[20]. Taq errors were identified by manual reconstruction of 16S rRNA secondary structures as pioneered in ref. 26 and detailed in ref. 20. In brief, the method scores as Taq errors sequence changes that violate either the sequence-conservation rule (nucleotides that are different in positions more than 98% conserved in all bacterial sequences) or the secondary-structure-conservation rule (apparent changes resulting in mismatches in stem structures that are not detected in related sequences). The Taq error rate determined from these rules was 3.3×10^{-5} per nucleotide per duplication, which agrees remarkably well with the experimentally determined value of 2×10^{-5} per nucleotide per duplication for the Taq polymerase used[20]. Further confidence that the large majority of Taq errors are captured is lent by several simple considerations. First, inspection of alignments showed the remaining variation clustered in regions known to be highly variable, which is inconsistent with the expected random distribution of Taq errors. Second, separate quantification of Taq error rate for positions falling under the conservation and the secondary structure rule previously gave highly similar rates of 1.7×10^{-5} and 2.5×10^{-5}, respectively[20]. Third, after 18 cycles, the inferred Taq error rate would lead to a misincorporation of bases at a rate of 3.6×10^{-4} per nucleotide. Because about 800 base pairs of sequence reads were obtained, this would translate into an average of 0.3 errors per sequence. This is close to the fraction of sequences (0.26; 181 of 686) removed from the modified library owing to identification of Taq errors. Last, potentially undetectable Taq errors by the secondary structure rule are those that change one allowed base pairing into another (for example, A-U to G-U). Although this can happen in two-thirds of all positions in rRNA, only 30% of the time will random replacement of one nucleotide by another due to Taq error result in another allowed base pairing. Therefore, at worst 20% (0.67×0.3) of Taq errors are missed but since only about 64% of the nucleotide positions fell under the secondary-structure rule this number translates into about 13%. Considering the probable incidence of errors per sequence based on a Taq error rate of 2×10^{-5} the calculation again results in an estimated 4% of sequences (0.13×0.3) that carry errors missed by the applied corrections.

The corrected set of sequences was used to estimate total diversity in the bacterioplankton community and patterns of relationships. An algorithm was developed[20] to group sequences into percentage similarity clusters (100%, 99%, 98%, and so on). This formed the basis for statistical extrapolation of total sequence diversity with the Chao-1 (ref. 7) and N_T/N_{MAX} (ref. 2) estimators. Accumulation of lineages through time was calculated with GENIE[27] from a non-optimized tree inferred by maximum likelihood with the molecular clock assumption enforced[8]. Identification of phylogenetic affiliation of the sequences was performed with the neighbour-joining method implemented in ARB[28] and followed by an analysis of more restricted groups of sequences by using distance and parsimony methods in PAUP* (ref. 29).

Received 10 February; accepted 11 May 2004; doi:10.1038/nature02649.

1. Rappé, M. S. & Giovannoni, S. J. The uncultured microbial majority. *Annu. Rev. Microbiol.* **57,** 369–394 (2003).
2. Curtis, T. P., Sloan, W. T. & Scannell, J. W. Estimating prokaryotic diversity and its limits. *Proc. Natl Acad. Sci. USA* **99,** 10494–10499 (2002).
3. Rosselló-Mora, R. & Amann, R. The species concept for prokaryotes. *FEMS Microbiol. Rev.* **25,** 39–67 (2001).
4. Cohan, F. M. What are bacterial species. *Annu. Rev. Microbiol.* **56,** 457–487 (2002).
5. Palys, T., Nakamura, L. K. & Cohan, F. M. Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int. J. Syst. Bacteriol.* **47,** 1145–1156 (1997).
6. Venter, C. J. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304,** 66–74 (2004).
7. Hughes, J. B., Hellmann, J. J., Ricketts, T. H. & Bohannan, B. J. M. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* **67,** 4399–4406 (2001).
8. Martin, A. P. Phylogenetic approaches for describing and comparing the diversity of microbial communities. *Appl. Environ. Microbiol.* **68,** 3673–3682 (2002).
9. Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V. & Polz, M. F. Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *J. Bacteriol.* **186,** 2629–2635 (2004).
10. Klappenbach, J. A., Dunbar, J. M. & Schmidt, T. M. rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* **66,** 1328–1333 (2000).
11. Hacker, J. & Carniel, E. Ecological fitness, genomic islands and bacterial pathogenicity—a Darwinian view of the evolution of microbes. *EMBO Rep.* **2,** 376–381 (2001).
12. Lan, R. T. & Reeves, P. R. When does a clone deserve a name? A perspective on bacterial species based on population genetics. *Trends Microbiol.* **9,** 419–424 (2001).
13. Welch, R. A. *et al.* Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. *Proc. Natl Acad. Sci. USA* **99,** 17020–17024 (2002).
14. Daubin, V., Moran, N. A. & Ochman, H. Phylogenetics and the cohesion of bacterial genomes. *Science* **301,** 829–832 (2003).
15. Rocap, G. *et al.* Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424,** 1042–1047 (2003).
16. Schleper, C. A. *et al.* Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Crenarchaeum symbiosum*. *J. Bacteriol.* **180,** 5003–5009 (1998).
17. Béja, O. *et al.* Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl. Environ. Microbiol.* **68,** 335–345 (2002).
18. Thingstad, T. F. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol. Oceanogr.* **45,** 1320–1328 (2000).
19. Lan, R. T. & Reeves, P. R. Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends Microbiol.* **8,** 396–401 (2000).
20. Klepac-Ceraj, V. *et al.* High overall diversity and dominance of microdiverse relationships in salt marsh sulfate-reducing bacteria. *Environ. Microbiol.* doi: 10.1111/j.1462–2920.2004.00600.x (2004).
21. Connon, S. A. & Giovannoni, S. J. High-throughput method for culturing microorganisms in very low nutrient media yield diverse new marine isolates. *Appl. Environ. Microbiol.* **768,** 3878–3885 (2002).
22. Vergin, K. L. *et al.* Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order *Planctomycetales*. *Appl. Environ. Microbiol.* **64,** 3075–3078 (1998).
23. Thompson, J. R., Marcelino, L. A. & Polz, M. F. Heteroduplexes in mixed-template amplifications: formation, consequences and elimination by 'reconditioning PCR'. *Nucleic Acids Res.* **30,** 2083–2088 (2002).
24. Cole, J. R. *et al.* The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res.* **31,** 442–443 (2003).
25. Hugenholtz, P. & Huber, T. Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. *Int. J. Syst. Evol. Microbiol.* **53,** 289–293 (2003).
26. Field, K. G. *et al.* Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl. Environ. Microbiol.* **63,** 63–70 (1997).
27. Pybus, O. G. & Rambaut, A. GENIE: Estimating demographic history from molecular phylogenies. *Bioinformatics* **18,** 1404–1405 (2003).
28. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32,** 1363–1371 (2004).
29. Swofford, D. L. *PAUP*. Phylogenetic Analysis Using Parsimony (*And Other Methods) Version 4* (Sinauer Associates, Sunderland, Massachusetts, 2002).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank S. Bertilsson (Uppsala University) for sample collection, D. Veneziano (MIT) for discussions on statistical extrapolation of sequence diversity, P. Chisholm (MIT) for numerous suggestions, and researchers and staff of the PIE-LTER (Plum Island Ecosystem-Long Term Ecological Research) for logistical support. This work was supported by grants from the National Science Foundation, the Department of Energy Genomes to Life program and a postdoctoral fellowship from the Spanish Ministry of Education to S.G.A.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.F.P. (mpolz@mit.edu). The sequences of the cloned partial 16S rRNA genes have been deposited in the GenBank database under accession numbers AY580336–AY580882.

Cambrian origins and affinities of an enigmatic fossil group of arthropods

N. E. Vaccari[1], G. D. Edgecombe[2] & C. Escudero[3]

[1]*Instituto de Geología y Minería, Universidad Nacional de Jujuy, Avenida Bolivia 1661, San Salvador de Jujuy (4600), Argentina*
[2]*Australian Museum, 6 College Street, Sydney, NSW 2010, Australia*
[3]*O'Higgins y Zarate, 108, Palpalá (4612), Jujuy, Argentina*

Euthycarcinoids are one of the most enigmatic arthropod groups, having been assigned to nearly all major clades of Arthropoda. Recent work has endorsed closest relationships with crustaceans[1] or a myriapod–hexapod assemblage[2], a basal position in the Euarthropoda[3], or a placement in the Hexapoda[4] or hexapod stem group[5]. Euthycarcinoids are known from 13 species ranging in age from Late Ordovician or Early Silurian to Middle Triassic, all in freshwater or brackish water environments[6]. Here we describe a euthycarcinoid from marine strata in Argentina dating from the latest Cambrian period, extending the group's record back as much as 50 million years. Despite its antiquity and marine occurrence, the Cambrian species demonstrates that morphological details were conserved in the transition to fresh water. Trackways in the same unit as the euthycarcinoid strengthen arguments that similar traces of subaerial origin from Cambro-Ordovician rocks were made by euthycarcinoids[7,8]. Large mandibles in euthycarcinoids[6,9] are confirmed by the Cambrian species. A morphology-based phylogeny resolves euthycarcinoids as stem-group Mandibulata, sister to the Myriapoda and Crustacea plus Hexapoda.

Mandibulata Snodgrass, 1938
Euthycarcinoidea Gall and Grauvogel, 1964
Euthycarciniformes Starobogatov, 1988

Apankura gen. nov.

Etymology. *Apankura* (Quechua), meaning crab.
Type species. *Apankura machu* gen. et sp. nov.
Diagnosis. Euthycarciniform with large mandibles that occupy most of the space beneath the posterior cephalic tergite; anterior two pairs of pre-abdominal limbs smaller than the posterior nine pairs; limbs markedly taper distally, composed of about ten podomeres, distal podomeres are shorter, large setae are absent; at least six post-abdominal segments; post-abdominal tergites are each about 2.5-times wider than they are long.

Apankura machu sp. nov.

Etymology. Genus as above; *machu* (Quechua), meaning grandfather.
Holotype. Museo de Geología, Mineralogía y Paleontología, Universidad Nacional de Jujuy (JUY-P 24; Fig. 1).
Locality and horizon. Bed of Río Huasamayo, Garganta del Diablo, near Tilcara, Jujuy Province, Argentina. The holotype (the only known specimen) is in greenish-grey mudstone from the Casa Colorada Member, Santa Rosita Formation. The trilobites *Neoparabolina frequens argentina* and *Plicatolina scalpta* on the same slab indicate a latest Cambrian age (lower part of *Neoparabolina frequens argentina* zone)[10]. Green shales of the Casa Colorada Member represent lower offshore deposition in an open marine facies[11].
Diagnosis. As for genus.

The holotype is 38 mm long, including the head, pre-abdomen and six segments of the post-abdomen. The maximum width of the pre-abdominal tergites is 16 mm. As in other euthycarcinoids[2,12], the head is composed of a short anterior tergite and a longer, wider posterior tergite. The latter is trapezoidal, with gently curved lateral margins. The antenna is uniramous, with at least nine short articles. Large, well-defined spheroidal processes[13] are at the lateral margin